Novel copper-containing membrane monooxygenases (CuMMOs) encoded by alkane-utilizing *Betaproteobacteria*

Running title: Novel CuMMOs in *Betaproteobacteria*

Fauziah F. Rochman[1], Miye Kwon[2], Roshan Khadka[1], Ivica Tamas[1,3], Azriel Abraham Lopez-Jauregui[1,4], Andriy Sheremet[1], Angela Smirnova[1], Rex R. Malmstrom[5], Sukhwan Yoon[2], Tanja Woyke[5], Peter F. Dunfield[1]*, Tobin J. Verbeke[1]

[1] Department of Biological Sciences, University of Calgary, 2500 University Dr. NW Calgary AB Canada T2N 1N4

[2] Department of Civil and Environmental Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Korea

[3]Department of Biology and Ecology, University of Novi Sad Novi Sad, Serbia

[4] Instituto Tecnologico y de Estudios Superiores de Monterrey, Chihuahua, Mexico

[5] Department of Energy Joint Genome Institute, Walnut Creek, California, USA.

*Corresponding author: pfdunfie@ucalgary.ca; 403-220-2469

Competing interests: The authors declare that they have no competing interests.

Abstract

Copper-containing membrane monooxygenases (CuMMOs) are encoded by *xmoCAB(D)* gene clusters and catalyze the oxidation of methane, ammonia, or some short chain alkanes and alkenes. In a metagenome constructed from an oilsands tailings pond we detected an *xmoCABD* gene cluster with <59% derived amino acid identity to genes from known bacteria. Stable isotope probing experiments combined with a specific *xmoA* qPCR assay demonstrated that the bacteria possessing these genes were incapable of methane assimilation, but did grow on ethane and propane. Single-cell genomes (SAGs) from propane-enriched samples were therefore constructed and screened with the specific PCR assay to identify bacteria possessing the target gene cluster. Multiple SAGs of *Betaproteobacteria* belonging to the genera *Rhodoferax* and *Polaromonas* possessed close homologues of the metagenomic *xmoCABD* gene cluster. Unexpectedly, each of these two genera also possessed other *xmoCABD* paralogs, representing two additional lineages in phylogenetic analyses. Metabolic reconstructions from SAGs predicted that neither bacterium was capable of catabolic methane or ammonia oxidation, but that both were capable of higher *n*-alkane degradation. The involvement of the encoded CuMMOs in alkane oxidation was further suggested by reverse transcription PCR analyses, which detected elevated transcription of the *xmoA* genes upon enrichment of water samples with propane as the sole energy source. Enrichments, isotope incorporation studies, genome reconstructions, and gene expression studies therefore all

agreed that the unknown *xmoCABD* operons did not encode methane or ammonia monooxygenases, but rather n-alkane monooxygenases. This study broadens the known diversity of CuMMOs and identifies non-nitrifying *Betaproteobacteria* as possessing these enzymes.

Introduction

Enzymes in the copper-containing membrane monooxygenase family (CuMMOs) catalyze diverse reactions. Particularly important are CuMMOs that act as ammonia and methane monooxygenases, as these play important roles in global carbon and nitrogen cycles (1-4). Nitrifying *Bacteria* and *Archaea* use ammonia monooxygenase (AMO) to catalyze the oxidation of ammonium to hydroxylamine, while methanotrophic bacteria use particulate methane monooxygenase (pMMO) to convert methane into methanol. Evidence for CuMMO-mediated metabolism of other compounds like short-chain alkanes and alkenes has emerged in recent years. CuMMOs have been reported to catalyse C2-C4 alkane oxidation in a number of actinobacterial strains including *Mycobacterium chubuense* NBB4 (5), *Mycobacterium rhodesiae* NBB3 (6) and *Nocardiodes* sp. CF8 (7). Ethylene-assimilating *Haliea* spp. within the class *Gammaproteobacteria* have also been shown to possess CuMMOs, although the role of the enzymes in ethylene oxidation is not firmly established (8). CuMMOs are also known to oxidise numerous other substrates via competitive co-oxidation, particularly hydrocarbons containing methyl and alkyl groups, but the converted substrates do not support growth (3).

CuMMOs are encoded in an operon of three to four genes in the usual order CAB(D). The D gene is sometimes encoded separate from the operon, or can be absent entirely (9). Operons are by convention named *amoCAB(D)* (encoding AMO), *pmoCAB(D)* (encoding pMMO), or other names depending

on the substrate-specificity. However, all are homologous and for simplicity we will refer to them collectively as *xmoCAB(D)* (10). Functionally and taxonomically coherent groups of methane and ammonia oxidisers are distinguishable on the basis of *xmoA* phylogeny, making these genes excellent biomarkers to identify environmental populations (10-12). Broad-spectrum PCR primers targeting these genes are therefore extensively used in ecological studies of methanotroph and nitrifier diversity. However, a common result of such studies is the detection of divergent sequences of unknown affiliation or function (12). Genome and metagenome studies are also uncovering new operons encoding divergent CuMMOs. Notable examples include the three different *xmoCAB* operons reported in verrucomicrobial methanotrophs (13-15) as well as the "*pxm*-group" identified in some alphaproteobacterial and gammaproteobacterial methanotrophs (16-18). Other divergent operons have been identified in sequenced genomes, including those of the gammaproteobacterium *Solimonas aquatica* DSM 25927 (19) and the betaproteobacterium *Hydrogenophaga* sp. T4 (Genbank accession number: AZSO00000000), but the functions of the encoded CuMMOs are unknown. Collectively, there is increasing evidence that the diversity of bacteria encoding CuMMO enzymes, and the diversity of substrates these enzymes act on, may be greater than currently appreciated.

Petroleum-impacted environments are good habitats to explore for new hydrocarbon degrading oxygenases. In the Athabascan oil sands of

Alberta, Canada, surface oil extraction involves a combination of alkali-hot water treatment and addition of chemical diluents (naphtha). The extraction process generates fluid tailings that are stored in open ponds to allow for particle settling, surface water recycling, and long-term pollutant containment (20). These tailings ponds contain high (up to 10 mM) concentrations of ammonia/ammonium (21, 22), along with residual bitumen-derived and naphtha-derived hydrocarbons including C_3-C_{14} alkanes and monoaromatics (benzene, toluene, ethylbenzene, xylene) (23, 24). Some oilsands tailings ponds are strongly methanogenic, and emit methane along with traces of other C_2-C_{10} volatile organic compounds (23-26). Aerobic methanotrophs possessing pMMO are abundant in oxic surface waters of these ponds (22).

Given the wealth of known CuMMOsubstrates in these tailings ponds, the oxic surface waters may offer a unique environment in which to discover new CuMMOs. Numerous investigations of the microbial communities in oil sands tailings ponds have been undertaken (20, 27), including metagenomic analyses (27-29). Through data mining of these metagenomes, a CuMMO-encoding operon highly divergent from any previously recognized operon was discovered. The objective of this study was to identify the bacteria encoding this sequence, and to gain insights into their metabolism.

Materials and methods

Sample sites and metagenomes

A metagenome (IMG Genome ID: 3300002856) of the surface oxic water of an active oilsands tailings pond near Fort McMurray, Alberta (West-In Pit or WIP) was generated on Illumina and Roche 454 platforms and assembled as described previously (29). An unusual *xmoCABD* operon (draft_100068512-draft_100068515) was identified in this metagenome and is referred to in this manuscript as "WIPMG *xmoCABD1*." The WIP pond was decommissioned in 2012 and repurposed as an End-Pit Lake (Base Mine Lake or BML) that no longer receives fresh tailings material (30). Therefore fresh samples could not be obtained from WIP for the present study, and were instead obtained from another active pond, Mildred Lake Settling Basin (MLSB). Until 2012 water was cycled between the two ponds and their microbial communities were very similar (22, 29), so we expected that MLSB would be a suitable proxy for the pre-2012 WIP community. Samples were obtained from the surface 10-cm of MLSB at several times between 2015-2017. The pond locations, chemical compositions, and the sampling methods used have been described previously (22, 29).

Bacterial enrichments

Surface water samples (0-10 cm depth) of MLSB sampled in August 2015 were used for enrichments and stable isotope probing (SIP) experiments. Twenty milliliter amounts were added to 100-ml serum bottles sealed with butyl rubber stoppers. Headspaces of triplicate capped bottles were augmented with 10% v/v methane, ethane, or propane. Alternatively,

ammonium chloride (20 mM) was added to enrich for nitrifying bacteria. The headspace of each bottle was supplemented with 5% v/v CO_2 to support autotrophy or anapleurotic CO_2 fixation. Bottles were incubated at 23°C with shaking (180 rpm) for up to 6 weeks. Gaseous hydrocarbon consumption and CO_2 production were determined using a Varian 450-gas chromatograph (Varian, Walnut Creek, CA) equipped with a thermal conductivity detector (detector T 150°) after separation in a 2 mm × 0.5 m Hayesep N column and a 2 mm × 1.2 m molecular sieve 16X column in series (column T 70°).

Detection and quantification of the WIPMG *xmoA1* gene

Water samples were centrifuged for 10 min at 10,000 × *g* prior to DNA extraction using the FastDNA Spin Kit for Soil (MP Biomedical, Santa Ana, CA). Eluted DNA was stored at -80°C. Specific PCR primers to target the WIPMG *xmoA1* gene were designed using the “Probe Design” tool in ARB (31) on a curated database of *xmoA* genes from public domain genomes. Primers and PCR assay conditions are detailed in Supplemental Table S1. Primer specificity was verified using Primer Blast (32) with 8 maximum allowed mismatches and the largest E-value of 10^5 against the NR database, including uncultured and environmental sample sequences. No unspecific hits were found. The PCR was optimized via temperature-gradient analysis, and reaction specificity verified via Sanger sequencing of selected PCR products, and melt curve analyses during qPCR. A PCR product was cloned into the vector pJET1.2 using the CloneJET PCR cloning kit (Thermo Fisher

Scientific, Waltham, MA), transformed into *E. coli*, recovered via colony PCR and used to construct qPCR standards ranging from 10^2 – 10^8 gene copies per microliter (33). qPCR was performed on a Qiagen RotorGene-Q (Qiagen, Toronto, ON) using SsoAdvanced Universal SYBR green supermix (Bio-Rad, Hercules, CA).

DNA-Stable isotope probing (SIP)

One-liter bottles containing 150 mL of MLSB water were sealed using butyl rubber stoppers and the headspace supplemented with 10% v/v of isotopically light (^{12}C) or heavy (99 mol% ^{13}C, Sigma-Aldrich, Oakville, Canada) methane, ethane, propane, or no added alkane (duplicate bottles of each). Five percent (v/v) $^{12}CO_2$ was also added to minimize cross-feeding of $^{13}CO_2$. Bottles were incubated as described above and gas depletion measured via GC. Experiments were stopped after ten days when between 21-33% of the supplied alkanes had been consumed. Extracted DNA was separated via isopycnic ultracentrifugation in cesium chloride and divided into twelve fractions of ~0.4 mL each, as described previously (34). The density of each fraction was measured using an AR200 refractometer (Reichert Technologies, Depew, NY). Recovered DNA was precipitated with polyethylene glycol and glycogen, washed with 70% ethanol, eluted, and quantified using the Quant-iT dsDNA HS assay kit (Invitrogen) (34). Samples from the SIP density fractions, unamended controls, and the initial (t=0)

community were investigated via qPCR of *xmoA1* genes, as well as via Illumina sequencing of 16S rRNA gene amplicons.

For amplicon sequencing multiple DNA density fractions of 1.72-1.74 g ml^{-1} were pooled to form a single representative ‘heavy DNA pool’. Fractions were selected if they contained much more DNA in ^{13}C versus ^{12}C incubations. Two controls were used: unfractionated DNA from the initial community; and the heaviest PCR-amplifiable fractions (1.71-1.73 g ml^{-1}) of the unamended samples. The latter control verified that designated "heavy" fractions were not simply GC-rich organisms. Amplification of the 341-785 region of 16S rRNA genes (35), and amplicon sequencing using an Illumina MiSeq (Illumina, San Diego, CA) was carried out as described previously (36). Reads were paired, filtered to exclude samples with quality-scores below 19 and analyzed using QIIME (37) with parameter settings described previously (29) Taxonomic identities were assigned via BLAST comparison to the Silva database (v. 123) (38). OTUs representing >1% of any relative read-set were validated through manual BLAST against the NCBI NR database.

Single-cell genomics

MLSB water sampled in September 2016 was enriched under 10% propane and 5% CO_2 as described above. Propane consumption was monitored using an SRI-8610C gas chromatograph (SRI Instruments, Torrance, CA) containing a HayeSep-D column (column T 190°C) coupled to a flame ionization detector (detector T 300°C) using N_2 as the carrier gas. When propane

consumption slowed, bottle headspaces were flushed with air and reconstituted with propane and CO_2. After 6 weeks total, 2-ml aliquots were removed and centrifuged at 300 × *g* for 2 min to remove inorganic particulate matter. The supernatant was transferred and cell biomass recovered via centrifugation at 6 000 × *g* for 3 mins. Cell pellets were washed three times in 50% strength PBS, then resuspended in 1 ml of 50 mM Tris-EDTA buffer (pH 8.0) containing 10% v/v glycerol. The prepared cells were then sorted into 384-well plates and single amplified genomes (SAGs) prepared using methods described previously (39, 40). SAGs were screened for 16S rRNA genes with standard protocols, and each well containing an identified 16S rRNA gene was then screened via the specific WIPMG *xmoA1* PCR assay. Ten SAGs positive for both 16S rRNA and WIPMG *xmoA1* were selected for complete genome sequencing on an Illumina NextSeq (40), followed by assembly and annotation using the standard operating procedure of the Joint Genome Institute's microbial annotation pipeline (41). Genome completeness and contamination for individual and combined SAGs were estimated using CheckM (42).

Comparative phylogenetics and DNA-DNA hybridizations

Phylogenetic analysis was performed on concatenated derived amino acid sequences of the three operonic *xmoCAB* genes. Sequences from publicly available genomes/metagenomes and sequences determined in this study

were aligned via Clustal Omega (43) and the tree constructed using maximum likelihood employing the LG model in Seaview 4.4.12 (44).

In silico DNA-DNA hybridizations were performed using the online Genome-to-Genome Distance Calculator v. 2.1 (45). Fasta nucleic-acid (.fna) files containing all assembled scaffolds for a specific SAG were compared against other SAGs within the genus in a pairwise fashion. Values were calculated by determining the sum of all identities found in high-scoring segment pairs divided by the overall high-scoring segment pair length (Formula 2 in the program) as recommended for incomplete, draft genomes (45).

Analysis of *xmoA* transcripts in propane-fed enrichment culture

To our knowledge, successful cloning and expression of CuMMO-encoding genes has been achieved only once in any model organism, for the butane MMO of a *Mycobacterium* (6). Therefore a cloning approach was considered unlikely to succeed, and we instead applied RT-qPCR to address the function of the WIPMG *xmoCABD1*. A sealed 2-L Duran glass bottle containing 1.5 L of mineral salts M10 medium (46) was inoculated with 75 mL of MLSB water that had been pre-enriched with propane as described above. The reactor was fed with a flow-through of mixed gas (propane and air at a ratio of 1:12) at a flow rate of 2.6 mL min^{-1}. The fed-batch reactor was maintained at 30 ºC in the dark and the aqueous phase stirred at 250 rpm. Cell density reached

257 OD_{600nm} of 0.171 after 96 h of incubation, after which the propane was shut
258 off and gas feed continued with air alone for another 24 h.

259 At intervals (0, 48, 96 and 120 h), 0.5-mL samples were taken,
260 immediately treated with 1 mL RNAprotect bacteria reagent (Qiagen), and
261 centrifuged at 5000 × *g* for 10 mins for RNA extraction. Parallel samples for
262 DNA extraction were prepared without the RNAprotect treatment. Cell pellets
263 were stored at -80ºC until analyses. Genomic DNA and the total RNA were
264 extracted using the DNeasy PowerSoil Kit (Qiagen) and RNeasy Mini Kit
265 (Qiagen), respectively. Three separate samples were processed in parallel to
266 ensure reproducibility. Before lysing the cells for the extraction of the total
267 RNA, one microliter of luciferase mRNA solution (Promega, Madison, WI)
268 diluted to 10^{10} copies μL^{-1} was added to each RNA extraction vial to account
269 for the RNA loss during the extraction and purification procedures (47). The
270 extracted RNA samples were treated with DNase I (Qiagen) and purified with
271 RNeasy MinElute Cleanup Kit (Qiagen) as previously described (48). The
272 purified total RNA samples were reverse-transcribed using SuperScript III
273 reverse transcriptase (Invitrogen).

274 The WIPMG *xmoA1* genes in the DNA and cDNA samples were
275 quantified via qPCR using the primer set TP2f + TP2r (Table S1). The primer
276 set and assay were redesigned from the xmoA1 assay described earlier in
277 order to universally target the entire cluster of related *xmoA1* genes found in
278 SAGs and metagenome sequences. Specific assays were also designed to
279 target the *xmoA2* gene from the *Rhodoferax* SAGs and the *xmoA2* gene from

the *Polaromonas* SAGs (Table S1). Primer specificity was verified as described earlier.

Results

Sequence discovery and phylogenetic analyses

Analysis of a previously published metagenome (29) identified a scaffold of 64485 bp that encoded a four-gene cluster homologous to known CuMMO encoding operons. Like most known pMMO- and AMO-encoding operons the genes were organized in the C-A-B order, with an additional *xmoD* gene. Phylogenetic analyses (Figure 1) showed that the *xmoCAB* (designated as WIPMG *xmoCAB1*) is most closely related to an operon in S*olimonas aquatica* DSM 25927, a gammaproteobacterium isolated from a freshwater spring in Taiwan (49). However, the individual CAB subunits share only 54%-66% derived amino acid identity with this strain (59% overall). Only one other gene was annotated on the genomic scaffold, a long-chain fatty acid transport protein showing a maximal amino acid identity of 67% to a protein in the alkane-oxidizing betaproteobacterium, *Thauera butanivorans* (50).

Enrichment and stable isotope probing with potential CuMMO substrates

Given the high sequence divergence of the WIPMG *xmoCABD1* relative to sequences from known methanotrophs and nitrifiers (Figure 1, Table S2-S3), we sought to identify a possible ecological role for the organism(s)

possessing it. A specific qPCR assay targeting the WIPMG *xmoA1* gene was used to analyse tailings pond water samples enriched with methane, ammonium, ethane, or propane. The number of WIPMG *xmoA1* gene copies was low (12,849 ± 2628 gene copies ml^{-1}) at the onset of the experiment and stayed relatively constant over six weeks of incubation in controls and in methane or ammonium enrichments (Figure 2). The only clearly stimulatory treatment was propane, where gene counts increased by over an order of magnitude, although ethane may have caused a small, transient increase. Enrichments with benzene, toluene, xylene or naphthalene (5 ml L^{-1}) for 4 weeks also showed no increase in WIPMG *xmoA1* gene counts (data not shown).

Water samples were then enriched using isotopically light (^{12}C) or heavy (^{13}C) methane, ethane and propane. Rapid oxidation was observed using both the ^{12}C and ^{13}C alkanes, showing maximal potential oxidation rates of 117, 90, and 63 µmol L^{-1} d^{-1} for methane, ethane and propane, respectively (Figure S1). In the density gradient-fractionated DNA from an unamended control sample, as well as in all enrichments using ^{12}C substrates, the peak amount of total DNA and the peak number of WIPMG *xmoA1* copies were detected in a DNA fraction of 1.69-1.70 g ml^{-1} (Figure 3). This therefore represented the natural peak density of the DNA from the entire community, and from the organisms containing the *xmoA1* gene.

In the ^{13}C-methane-incubated samples the peak amount of DNA shifted to 1.71-1.74 g ml^{-1} due to incorporation of the heavy-isotope by

methanotrophs. However, there was no shift in the fraction containing the highest WIPMG *xmoA1* gene copy number due to ^{13}C methane assimilation. In both the ^{13}C and ^{12}C methane enrichments, peak WIPMG *xmoA1* copies were observed in light (1.69 g ml^{-1}) fractions, suggesting that the bacteria possessing these genes did not assimilate methane-derived carbon (Figures 3A, 3B).

In contrast, the density fraction showing the maximum WIPMG *xmoA1* copy numbers did shift after ^{13}C-ethane and ^{13}C-propane enrichment (Figures 3D, 3F). In both cases, WIPMG *xmoA1* gene copy numbers were highest at densities of 1.69-1.70 g ml^{-1} in the ^{12}C enrichments but shifted to >1.71 g ml^{-1} after enrichment with ^{13}C-labeled n-alkanes (Figures 3C, 3E). This shift suggests that the bacteria encoding WIPMG *xmoCAB1* were capable of assimilating carbon from ethane and propane. Peak WIPMG *xmoA1* copy numbers were three orders of magnitude higher in the propane enrichment, suggesting that this substrate was preferred over ethane.

16S rRNA gene amplicons of DNA from control samples (i.e. the entire unenriched tailings pond DNA sample, as well as just the heaviest PCR-amplifiable DNA fraction) showed diverse communities with 707-1098 OTUs detected. The heavy DNA fractions from the methane, ethane, and propane enrichments were dominated by fewer OTUs (367-622). *Gammaproteobacteria* was the predominant class in the unamended sample as well as the methane enrichment (Figure S2). The methane enrichment was dominated by a single OTU (Figure S3) closely related to the

methanotrophs *Methyloparacoccus* and *Methylocaldum* (97% sequence identity) agreeing with previous studies (22). *Betaproteobacteria* were much more abundant in the ethane and propane enrichments, comprising 48% and 77% of the total read sets respectively (Figure S2). Relative abundances of OTUs within the genera *Methyloversatilis, Hydrogenophaga*, *Pedomicrobium*, *Arenimonas, Acidovorax*, *Rhodoferax* and *Oxalicibacterium* increased after propane enrichment (Figure S3).

Screening single-cell genomes for *xmoCAB* genes.

About 98% of the identified sorted cells from a propane enrichment belonged to the class *Betaproteobacteria* and six distinct genera were identified overall (Table S4). Three of these, *Rhodoferax*, *Hydrogenophaga* and *Methyloversatilis*, had been shown to assimilate propane in SIP enrichments (Figure S3). However, the most abundant genus sorted was *Polaromonas*, which was not enriched in the propane SIP experiments. The water samples used for these two experiments were taken at different times, probably explaining the slight differences in the predominant bacteria.

Aliquots of amplified genomic DNA from the sorted wells were screened using the WIPMG *xmoA1*-specific PCR assay. Bands of the expected size were observed in multiple SAGs identified as *Rhodoferax* and *Polaromonas*, so five SAGs of each genus were selected for genome sequencing (Table S5). Comparative analyses suggested that the 5 SAGs of each genus were nearly clonal. *In silico* DNA-DNA hybridizations (45) of the

draft genomes within each genus were all >70% identical, suggesting each genus was represented by a single species in the sorted plates (Table S6, Table S7). The 16S rRNA gene sequences from the *Polaromonas* SAGs were identical and showed 98.0% nucleotide identity to *Polaromonas naphthalenivorans* CJ2, an aromatic hydrocarbon degrading bacterium (51). For the *Rhodoferax* genomes, the full length 16S rRNA gene sequences were identical except for a single nucleotide mismatch in SAG-1 and closely matched (98.5%) *Rhodoferax ferrireducens* T118 (52). Finished genomes for both *P. naphthalenivorans* CJ2 and *R. ferrireducens* T118 are available, but neither organism (nor any closely related genome-sequenced strain) possesses CuMMO-encoding genes.

Collectively, the *Rhodoferax* SAGs possessed two divergent *xmoCABD* operons (Figures 1, 4). One showed 99.9% nucleotide identity to the WIPMG *xmoCAB1* genes found in the original oilsands tailings pond metagenome. However, the second operon clustered in a distinct clade (Figure 1). The *Polaromonas* SAGs also encoded two divergent operons. Again, one was homologous to the WIPMG *xmoCAB1*. The other formed a third new lineage not homologous to the second operon in the *Rhodoferax* SAGs (Figure 1, Table S3). An orphan *xmoC* (e.g. Ga0215891_10812) was also identified in three of the five *Polaromonas* genomes, with flanking genes on the scaffolds confirming the absence of a complete operon.

The four new *xmo* gene clusters and selected neighboring genes are shown in Figure 4. Each of the four clusters included *xmoD*, a gene

occasionally part of an *xmoCAB(D)* operon in methanotrophs and nitrifiers, and occasionally present elsewhere in the genome (9). Promoter prediction with Virtual Footprint (53) indicated that these genes are expressed as single *xmoCABD* operon in all four cases. Other genes located nearby include genes encoding predicted alcohol and aldehyde dehydrogenases, which may be involved in degrading the downstream products of the monooxygenase reaction (Figure 4).

Metabolic potential

While the major goal of sequencing the SAGs was to identify the organisms containing WIPMG *xmoCABD1* operons, they were also analysed to indicate any potential for ammonia, methane, or alkane oxidation. The CheckM genome completeness estimates for the combined *Polaromonas* SAGs was 96% and for the combined *Rhodoferax* SAGs was 82% (Table S5). Contamination was estimated to be low (≤ 0.03%) or zero (Table S5). Therefore the combined SAGs should give nearly-complete overviews of metabolic capacity, especially for the *Polaromonas*.

A complete Calvin Benson Bassham (CBB) cycle for autotrophic CO_2 fixation, including the large subunit of ribulose bisphosphate carboxylase, was detected in the *Polaromonas* (Ga0215911_14316; Ga215901_1152) but not in the *Rhodoferax*. *Polaromonas* may therefore be capable of 1-C fixation via the CBB cycle, a prerequisite for autotrophic nitrification. However, hydroxylamine oxidoreductase, an enzyme essential for ammonia oxidation

418(54), was not annotated in any SAG. Each genus also possessed a diversity of
419genes encoding transporters and metabolic modules for organoheterotrophic
420growth, which would be very atypical for known bacterial nitrifiers (55).
421 Pathways typical of proteobacterial methanotrophs were also mostly
422missing, although some catabolism of 1-C substrates (formate and poossibly
423methanol) was encoded. Neither a ribulose monophosphate (RuMP) nor a
424serine cycle for fixation of 1-C intermediates of methane oxidation was
425complete in either organism. Genes encoding the key RuMP enzymes 3-
426hexulose-6-phosphate synthase and 6-phospho-3-hexuloisomerase were not
427found, nor were genes for the key serine cycle enzyme serine glyoxylate
428aminotransferase. A hydroxypyruvate reductase encoding gene was
429annotated, but was different from the form other methanotrophs use for the
430serine cycle (EC 1.1.1.81 instead of 1.1.1.29). There was no clear *mxaFI* or
431*xoxF*-encoded methanol dehydrogenase, which is common in methanotrophs
432(56, 57). However, two other pyrrolquinoline quinone (PQQ)-binding alcohol
433dehydrogenases were encoded in both *Rhodoferax* and *Polaromonas*
434genomes: a homologue to the single subunit mdh2-type methanol
435dehydrogenase identified in the methylotroph *Methyloversatilis universalis*
436FAM5 (58) (e.g. Ga0215885_1254 – *Rhodoferax*; Ga0215901_10418–
437*Polaromonas*), and a second PQQ-binding alcohol dehydrogenase (e.g.
438Ga0215904_10692 – *Rhodoferax*; Ga0215901_1397– *Polaromonas*) located
439just downstream of CuMMO-encoding subunits (Figure 4). Genes encoding
440PQQ synthesis were also identified in the *Polaromonas* (e.g

Ga0215902_10817-108110). Formaldehyde dehydrogenase or a tetrahydromethanopterin-linked pathway to convert formaldehyde to formate were not encoded, although other aldehyde dehydrogenases were present, often adjacent to the *xmoCABD* operons (Figure 4). The *Polaromonas* did encode multiple subunits of a formate dehydrogenase (e.g. Ga0215892_11102 to 11104).

On the other hand, complete pathways for propane oxidation were predicted in both *Rhodoferax* and *Polaromonas* genomes (Figure 5). In addition to the two CuMMOs, both genera encode multiple other alkane monooxygenases, including a group 3 (methane/alkane) soluble di-iron monooxygenase in the *Polaromonas* that is closely homologous (87% identity) to the butane monooxygenase of *Thauera butanivorans* (59, 60) (Table S8). The oxidation of propane could be initiated by one or several of these enzymes, either terminally forming 1-propanol or sub-terminally forming 2-propanol (61, 62). The terminal oxidation of 1-propanol could proceed via propionaldehyde and propionate (62) which could then be further oxidized by a number of described heterotrophic pathways (63). In brief, both organisms can theoretically convert 1-propanol to propionate and then propionyl-CoA (Figure 5). At this branch point, one possible degradation route includes oxidation via the citramalate cycle where propionyl-CoA is converted into succinyl-CoA (63, 64). In the *Rhodoferax*, the genes encoding a propionyl-CoA carboxylase (e.g. Ga0215895_1164-1165) and a methylmalonyl-CoA mutase (e.g. Ga0215895_1161) are adjacent in the

464genome. A similar gene neighbourhood architecture is observed in the
465*Polaromonas* genomes. Neither organism had gene homologues encoding
466known methylmalonyl-CoA epimerases, which catalyze the conversion
467between 2R-methylmalonyl-CoA and 2S-methylmalonyl-CoA (Figure 5).
468Multiple other epimerases are annotated in each genome, however, which
469may act as functional equivalents.

470 Another possible pathway includes the conversion of propionyl-CoA
471(plus oxaloacetate) to pyruvate (plus succinate) via the methylcitrate
472pathway (Figure 5A). In *Polaromonas*, genes encoding 2-methylcitrate
473synthase, 2-methylcitrate dehydratase and 2-methylisocitrate lyase are
474adjacent in the genome (e.g. Ga0215912_10019-100111). Genes for 2-
475methylcitrate synthase could not be identified in the *Rhodoferax* genomes
476although a citrate synthase encoding gene (e.g. Ga0215903_11011) is
477located just downstream of annotated 2-methylcitrate dehydratase and 2-
478methylcitrate lyase encoding genes.

479 A number of possible sub-terminal oxidation pathways have been
480described (62, 65). No known 2-propanol degradation pathways were
481identified in the *Rhodoferax* SAGs (Figure 5B), but in *Polaromonas*, three
482distinct NAD(P)-dependent alcohol dehydrogenase encoding genes (ADH) are
483adjacent to an annotated acetone/cyclohexanone monooxygenase (e.g.
484Ga0215909_1235). In the actinomycete *Gordonia* sp. TY-5, where the
485pathway of 2-propanol oxidation via acetone was first described (66), the
486acetone monooxygenase is adjacent to a methylacetate hydrolase (an

esterase). A homologous methylacetate hydrolase was not identified in the *Polaromonas* SAGs, but an esterase of unspecified activity is located just upstream of its acetone monooxygenase.

Transcription of *xmoA* in tailings pond water enriched with propane

The transcription of WIPMG *xmoA1* was quantified in an enrichment culture grown in a batch reactor continuously fed with propane (Table 1). The number of *xmoA1* genes increased from $8.0 \times 10^4 \pm 7.9 \times 10^3$ copies mL^{-1} immediately after inoculation (t = 0) to $3.9 \times 10^6 \pm 6.0 \times 10^5$ copies mL^{-1} at t = 96. Up-regulation of *xmoA1* transcription during growth on propane was indicated by significantly higher transcript-to-gene ratios ($p<0.05$) observed at t = 48 h and t = 96 h than at t = 120 h, 24 hours after shutdown of propane supply (Table 1).

The *xmoA2* gene found in *Polaromonas* was not detectable with a specific PCR assay, however the *xmoA2* gene in the *Rhodoferax* SAGs was. Fewer *Rhodoferax xmoA2* gene copies were detected compared to WIPMG *xmoA1* genes, suggesting that fewer bacteria in the enrichment had a close homologue of this *xmoA2* cluster (Table 1). However, like *xmoA1*, expression of the *Rhodoferax xmoA2* was also dependent on the propane supply, although not as dramatically. Expression of *xmoA2* decreased only 4-fold (rather than 30 fold for the *xmoA1*) after removal of propane. Although they were not the predominant bacteria in the enrichmment, 16S rRNA gene amplicon analysis detected the genera XX at low relative abundances (XX%

of XX total reads). Sequencing of the *xmoA1* gene product obtained from the enrichment culture also verified that it matched 100% to the gene from XXXX.

Discussion

Phylogenies of genes encoding CuMMOs can clearly delineate some functional and taxonomic groups of bacteria. Different ammonia oxidisers (in the phylum *Thaumarchaeota,* and the proteobacterial classes *Gammaproteobacteria*, and *Betaproteobacteria*), methane oxidisers (phyla *Verrucomicrobia* and NC10, proteobacterial classes *Gammaproteobacteria*, and *Alphaproteobacteria*), and butane oxidisers (phylum *Actinobacteria*) can all be reliably separated on the basis of phylogenetic clustering (10-12, 16). This coherent phylogenetic structure has served as a useful backbone to establishing community structure-function relationships in numerous molecular ecology surveys investigating methane and ammonium oxidisers (10, 12). However, phylogenetic clusters of CuMMO-encoding genes with unknown function and taxonomic affiliation are also found in genomes, metagenomes, and environmrntal PCR amplicons produced with broad-specificity PCR primers. Here we investigated one such unknown *xmoCABD* operon identified in the metagenome of an oilsands tailings pond, in order to assign it to a probable taxon and function. Single-cell genomics positively identified bacteria possessing this operon in our samples as members of the class *Betaproteobacteria*, in the family *Comamonadaceae* and the genera

533*Rhodoferax* and *Polaromonas. Comamonadaceae* are abundant in
534hydrocarbon-contaminated environments such as oilsands tailings, and
535specific members possess multiple aerobic and anaerobic petroleum-
536degrading pathways (20, 27, 29, 67). None are known to oxidise methane or
537ammonia.

538 The use of a single-cell genomics approach was preferred to
539metagenomic binning of genomes because nearly all *xmo* operons differ
540significantly in their nucleotide compositional biases compared to their
541overall host genomes (10), which could potentially cause problems with
542compositional binning. The assignment of the *xmo* operons to the
543*Rhodoferax* and *Polaromonas* genomes was verified in multiple
544uncontaminated SAGs of each bacterium. Unexpectedly, the SAGs also
545revealed other divergent CuMMO-encoding operons in both bacteria.
546Collectively, the CuMMOs clustered into three distinct clades. In each clade
547the closest genes from identified, cultured bacteria are only distantly related
548to those from our SAGs, and formal studies into gene expression or enzyme
549function of the *xmo* genes in these bacteria have not been reported. The
550closest genome sequence to the WIPMG *xmoCAB1* cluster showed only 59%
551amino acid (AA) identity and is encoded by the gammaproteobacterium
552*Solimonas aquatica* NAA16^{T}. *Solimonas aquatica* is a metabolically versatile
553bacterium (49) whose CuMMO-enzymes were identified solely through
554genome sequencing as part of the Genomic Encyclopedia of Archaeal and
555Bacterial Type Strains, Phase-II sequencing initiative (19). Characterization of

556gene expression and alkane metabolism in the type strain *Solimonas*
557*aquatica* NAA16^{T} could be valuable to explore the diversity of CuMMOs,
558however the low similarities of the *Solimonas xmo* operon to those of the
559*Polaromonas/Rhodoferax* characterized in this study suggests that the
560functions of the encoded CuMMOs may not be the same.

561 The low numbers of WIPMG *xmoA1* genes observed under ammonia
562enrichments (Figure 2) suggested that neither the *Rhodoferax* nor the
563*Polaromonas* strains were capable of growth via nitrification. This was further
564supported by the lack of a hydroxylamine oxidoreductase in any of the SAGs.
565Methane oxidation also seemed an unlikely function based on SIP and qPCR
566analyses after incubation under methane-containing atmospheres (Figure
5672,3). No enrichment of the *xmoA* or 16S rRNA genes of these bacteria was
568observed in the heavy fraction of the ^{13}C-methane SIP experiments, verifying
569these bacteria did not assimilate methane. The genomic compositions were
570also not typical of methanotrophs, although a limited capacity to catabolise
571some 1-C compounds was indicated. Neither bacterium encoded methanol
572and formaldehyde oxidation pathways typical of methanotrophs and neither
573encoded pathways for assimilating methane-carbon via formaldehyde or
574formate. The *Polaromonas* encoded some capability for 1-C catabolism
575(formate and possibly methanol), and although methane assimilation was
576not encoded, the CBB cycle would provide an alternative path of C
577assimilation, as practiced by *Verrucomicrobia* methanotrophs (34). Given this
578potential, and the known promiscuous nature of CuMMOs (3), methane

579cannot be completely discounted as a possible substrate for the new
580CuMMOs. It is possible that the *Polaromonas/Rhodoferax* obtain energy from
581the oxidation of methane, but are unable to assimilate it. However, a clear
582test of this hypothesis would require a pure culture. Our cultivation efforts to
583date have included plating and batch-culture dilution under propane-
584containing atmospheres, however the target organisms have proven elusive,
585and are easily overgrown by other species. Successful isolation will likely
586require optimizing incubations conditions such as pO_2.

587 Contrary to methane and ammonia oxidation, n-alkane metabolism in
588the target bacteria was supported by multiple lines of evidence: enrichment,
589SIP, genome analysis, and gene expression. Genomic analyses suggested
590two plausible pathways for terminal propane oxidation in both *Rhodoferax*
591and *Polaromonas* (Figure 4A) along with a pathway in *Polaromonas* for the
592oxidation of 2-propanol (Figure 4B). These genome predictions, along with
593the strong assimilation of propane-derived C into the *xmoA* genes of
594*Rhodoferax* and *Polaromonas* demonstrated by SIP studies, show clearly that
595these bacteria are capable of alkanotrophy. They do not together prove that
596the CuMMOs are key enzymes in propane oxidation, since multiple other
597hydrocarbon monooxygenases were also identified in the genomes (Table
598S8). However, RT-qPCR studies indicated that WIPMG *xmoA1* gene
599expression in an enrichment culture (as well as the expression of the *xmoA2*
600gene detected in the *Rhodoferax*) is regulated depending on the availability

601of propane, strongly suggesting the involvement of these CuMMO enzymes
602in propane oxygenation.

603 CuMMO-enabled n-alkane (butane, ethane, and propane) oxygenation
604has already been established in some *Actinomycetes* (5-7). Although our
605bacteria showed a clear preference for growth on propane, ethane also
606supported a lower growth and C-assimilation rate, and could also have been
607a CuMMO substrate. A difference between the CuMMO-encoding genes in our
608*Comomonadaceae* versus those of *Actinobacteria* is the presence of *xmoD* in
609the former. The product of *xmoD* was recently described as a critical
610component of some CuMMOs: a Cu-containing polypeptide that may
611facilitate assembly and stabilization of the CuMMO complex, or facilitate
612electron delivery to the active site (9). The gene is present in all bacteria
613encoding AMO or pMMO for ammonia/methane oxidation, sometimes as part
614of the *xmo* operon and sometimes elsewhere in the genome. However,
615homologues are not found in the genomes of the actinobacteria
616*Mycobacteria* and *Nocardiodes* that encode n-alkane targeting CuMMOs (9).
617This suggests that the *Rhodoferax/Polaromonas* CuMMOs identified in our
618study are functionally more similar to the better known proteobacterial
619pMMO/AMO enzymes than to the actinobacterial CuMMOs, a proposition
620supported by phylogenetic analysis, which places the actinobacterial
621CuMMOs well apart from all proteobacterial CuMMOs (Figure 1).

622 This study has expanded the known diversity of *xmoCAB(D)* operons
623encoding CuMMOs and the taxonomic groups that encode this enzyme.

Definitive functional roles for any of the encoded CuMMOs could only be inferred, and conclusive evidence will require further experimentation using laboratory cultures. However, multiple lines of cultivation-independent evidence suggest that these CuMMOs are probably involved primarily in n-alkane oxidation, rather than methane or ammonia oxidation.

Acknowledgements

This work was made possible through the NSERC (Natural Sciences and Engineering Research Council of Canada) Collaborative Research and Development program (Grant number CRDPJ478071-14); an NSERC Discovery grant (2014-05067); and a grant from the National Research Foundation of Korea (2018K1A3A1A74065626). Initial metagenomic data were created with financial assistance from Genome Canada, Genome Alberta, Genome BC and the Government of Alberta (GC Grant 1203). SAG sequencing and analysis was supported by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility supported under Contract No. DE-AC02-05CH11231. We acknowledge the assistance of Syncrude Canada, Ltd in providing samples. AA Lopez-Jauregui was supported by a Mitacs Globallink internship award.

Competing interests

The authors declare that they have no competing interests.

Supplementary information

Supplementary information is available at The ISME Journal's website.

651

652**References**X1. Hakemian AS, Rosenzweig AC. The biochemistry of
653methane oxidation. Ann Rev Biochem. 2007; **76**: 223-241.
6542. Canfield DE, Glazer AN, Falkowski PG. The evolution and future of
655Earth's nitrogen cycle. Science. 2010; **330**: 192-196.
6563. Semrau JD. Bioremediation via methanotrophy: Overview of recent
657findings and suggestions for future research. Frontier Microbiol. 2011; **2**:
658209.
6594. Bodelier PLE, Steenbergh AK. Interactions between methane and the
660nitrogen cycle in light of climate change. Curr Opinion Environ Sustainability.
6612014; **9**: 26-36.
6625. Coleman NV, Yau S, Wilson NL, Nolan LM, Migocki MD, Ly MA, et al.
663Untangling the multiple monooxygenases of *Mycobacterium chubuense*
664strain NBB4, a versatile hydrocarbon degrader. Environ Microbiol Rep. 2011;
665**3**: 297-307.
6666. Coleman NV, Le NB, Ly MA, Ogawa HE, McCarl V, Wilson NL, et al.
667Hydrocarbon monooxygenase in *Mycobacterium*: recombinant expression of
668a member of the ammonia monooxygenase superfamily. ISME J. 2012; **6**:
669171-182.
6707. Sayavedra-Soto LA, Hamamura N, Liu CW, Kimbrel JA, Chang JH, Arp DJ.
671The membrane-associated monooxygenase in the butane-oxidizing Gram-
672positive bacterium *Nocardioides* sp. strain CF8 is a novel member of the
673AMO/PMO family. Environ Microbiol Rep. 2011; **3**: 390-396.

6748. Suzuki T, Nakamura T, Fuse H. Isolation of two novel marine ethylene-
675assimilating bacteria, *Haliea* species ETY-M and ETY-NAG, containing
676particulate methane monooxygenase-like genes. Microbes Environment.
6772012; **27**: 54-60.
6789. Fisher OS, Kenney GE, Ross MO, Ro SY, Lemma BE, Batelu S, et al.
679Characterization of a long overlooked copper protein from methane- and
680ammonia-oxidizing bacteria. Nat Comm. 2018; **9**: 4276.
68110. Khadka R, Clothier L, Wang. L, Lim CK, Klotz MG, Dunfield PF.
682Evolutionary history of copper membrane monooxygenases. Frontier
683Microbiol. 2018; **9**: 2493.
68411. Purkhold U, Pommerening-Röser A, Juretschko S, Schmid MC, Koops HP,
685Wagner M. Phylogeny of all recognized species of ammonia oxidizers based
686on comparative 16S rRNA and *amoA* sequences analysis: implications for
687molecular diversity surveys. Appl Environ Microb. 2000; **66**: 5368-5382.
68812. Knief C. Diversity and habitat preferences of cultivated and
689uncultivated aerobic methanotrophic bacteria evaluated based on *pmoA* as
690molecular marker. Frontier Microbiol. 2015; 6: 1346.
69113. Op den Camp HJ, Islam T, Stott MB, Harhangi HR, Hynes A, Schouten S,
692et al. Environmental, genomic and taxonomic perspectives on
693methanotrophic Verrucomicrobia. Environ Microbiol Rep. 2009; **1**: 293-306.
69414. Khadem AF, van Teeseling MC, van Niftrik L, Jetten MS, Op den Camp
695HJ, Pol A. Genomic and physiological analysis of carbon storage in the

verrucomicrobial methanotroph "*Ca. Methylacidiphilum fumariolicum*" SolV. Frontier Microbiol. 2012; **3**: 345.

15. Anvar SY, Frank J, Pol A, Schmitz A, Kraaijeveld K, Dunnen JTd, et al. The genomic landscape of the verrucomicrobial methanotroph *Methacidiphilum fumariolicum* SolV. BMC Genomics. 2014; **15**: 914.

16. Tavormina PL, Orphan VJ, Kalyuzhnaya MG, Jetten MS, Klotz MG. A novel family of functional operons encoding methane/ammonia monooxygenase-related proteins in gammaproteobacterial methanotrophs. Environ Microbiol Report. 2011; **3**: 91-100.

17. Vorobev AV, Jagadevan S, Jain S, Anantharaman K, Dick GJ, Vuilleumier S, et al. Genomic and transcriptomic analyses of the facultative methantroph *Methylocystis* sp. Strain Sb2 grown on methane or ethanol. Appl Environ Microb. 2014; **80**: 3044-3052.

18. Kits KD, Klotz MG, Stein LY. Methane oxidation coupled to nitrate reduction under hypoxia by the Gammaproteobacterium *Methylomonas denitrificans*, sp. nov. type strain FJG1. Environ Microbiol. 2015; **17**: 3219-3232.

19. Whitman WB, Woyke T, Klenk HP, Zhou Y, Lilburn TG, Beck BJ, et al. Genomic Encyclopedia of Bacterial and Archaeal type strains, phase III: the genomes of soil and plant-associated and newly described type strains. Stand Genomic Sci. 2015; **10**: 26.

20. Foght JM, Gieg LM, Siddique T. The microbiology of oil sands tailings: past, present, future. FEMS Microbiol Ecol. 2017; **93**: 5.

71921. Allen EW. Process water treatment in Canada’s oil sands industry: I.
720Target pollutants and treatment objectives. J Environ Eng Sci. 2008; **7**: 123-
721138.

72222. Saidi-Mehrabad A, He Z, Tamas I, Sharp CE, Brady AL, Rochman FF, et
723al. Methanotrophic bacteria in oilsands tailings ponds of northern Alberta.
724ISME J. 2013; **7**: 908-921.

72523. Siddique T, Shahimin MFM, Zamir S, Semple K, Li C, Foght JM. Long-
726term incubation reveals methanogenic biodegradation of C5 and C6 iso-
727alkanes in oil sands tailings. Environ Sci Technol. 2015; **49**: 14732-14739.

72824. Siddique T, Fedorak PM, Mackinnon MD, Foght JM. Metabolism of BTEX
729and naphtha compounds to methane in oil sands tailings. Environ Sci
730Technol. 2007; **41**: 2350-2356.

73125. Simpson IJ, Blake NJ, Barletta B, Diskin GS, Fuelberg HE, Gorham K, et
732al. Characterization of trace gases measured over Alberta oil sands mining
733operations: 76 speciated C_2–C_{10} volatile organic compounds (VOCs), CO_2, CH_4,
734CO, NO, NO_2, NOy, O_3 and SO_2. Atmos Chem Phys. 2010; **10**: 11931-11954.

73526. Small CC, Cho S, Hashisho Z, Ulrich AC. Emission from oil sands tailings
736ponds: Review of tailings pond parameters and emission estimates. J
737Petroleum Sci Eng. 2015; **127**: 490-501.

73827. An D, Caffrey SM, Soh J, Agrawal A, Brown D, Budwill K, et al.
739Metagenomics of hydrocarbon resource environments indicates aerobic taxa
740and genes to be unexpectedly common. Environ Sci Technol. 2013; **47**:
74110708-10717.

74228. Aguilar M, Richardson E, Tan B, Walker G, Dunfield PF, Bass D, et al.
743Next-generation sequencing assessment of eukaryotic diversity in oil sands
744tailings ponds sediments and surface water. J Eukaryotic Microbiol. 2016; **63**:
745732-743.

74629. Rochman FF, Sheremet A, Tamas I, Saidi-Mehrabad A, Kim JJ, Dong X,
747et al. Benzene and naphthalene degrading bacterial communities in an oil
748sands tailings pond. Frontier Microbiol. 2017; **8**: 1845.

74930. Syncrude. "Mildred Lake extension project". Public Disclosure
750Document (Fort McMurray: Syncrude Canada Ltd) [Internet]. 2012. Available
751from: http://www.syncrude.ca/assets/pdf/News-Room/MLX-Project-Public-
752Disclosure-Document.pdf.

75331. Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar, et al.
754ARB: a software environment for sequence data. Nucleic Acid Res. 2004; 32:
7551363-1371.

75632. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL.
757Primer-BLAST: A tool to design target-specific primers for polymerase chain
758reaction. BMC Bioinformatics. 2012; **13**: 134.

75933. Sharp CE, Smirnova AV, Graham JM, Stott MB, Khadka R, Moore TR, et
760al. Distribution and diversity of Verrucomicrobia methanotrophs in
761geothermal and acidic environments. Environ Microbiol. 2014; **16**: 1867-
7621878.

76334. Sharp CE, Stott MB, Dunfield PF. Detection of autotrophic
764verrucomicrobial methanotrophs in a geothermal environment using stable
765isotope probing. Frontier Microbiol. 2012; **3**: 303.

76635. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, et al.
767Evaluation of general 16S ribosomal RNA gene PCR primers for classical and
768next-generation sequencing-based diversity studies. Nucleic Acid Res. 2013;
769**41**: e1.

77036. Ruhl IA, Grasby SE, Haupt ES, Dunfield PF. Analysis of microbial
771communities in natural halite springs reveals a strong but domain-dependent
772relationship of species diversity to osmotic stress. Environ Microbiol Rep.
7732018; **10**: 695-703.

77437. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD,
775Costello EK, et al. QIIME allows analysis of high-throughput community
776sequencing data. Nat Method. 2010; **7**: 335-336.

77738. Glockner FO, Yilmaz P, Quast C, Gerken J, Beccati A, Ciuprina A, et al.
77825 years of serving the community with ribosomal RNA gene reference
779databases and tools. J Biotechnol. 2017; **261**: 169-176.

78039. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng JF,
781et al. Insights into the phylogeny and coding potential of microbial dark
782matter. Nature. 2013; **499**: 431-437.

78340. Rinke C, Lee J, Nath N, Goudeau D, Thompson B, Poulton N, et al.
784Obtaining genomes from uncultivated environmental microorganisms using
785FACS-based single-cell genomics. Nature protocol. 2014; **9**: 1038-1048.

78641. Huntemann M, Ivanova NN, Mavromatis K, Tripp HJ, Paez-Espino D,
787Palaniappan K, et al. The standard operating procedure of the DOE-JGI
788Microbial Genome Annotation Pipeline (MGAP v.4). Stand Genomic Sci. 2015;
789**10**: 86.

79042. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW.
791Assessing the quality of microbial genomes recovered from isolates, single
792cells, and metagenomes. Genome Res. 2014; **25**: 1043-1055.

79343. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast,
794scalable generation of high-quality protein multiple sequence alignments
795using Clustal Omega. Mol System Biol. 2011; **7**: 539.

79644. Gouy M, Guindon S, Gascuel O. SeaView version 4: A multiplatform
797graphical user interface for sequence alignment and phylogenetic tree
798building. Mol Biol Evol. 2010; **27**: 221-224.

79945. Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. Genome sequence-
800based species delimitation with confidence intervals and improved distance
801functions. BMC Bioinformatics. 2013; **14**: 60.

80246. Heyer J, Galchenko VF, Dunfield PF. Molecular phylogeny of type II
803methane-oxidizing bacteria isolated from various environments. Microbiol.
8042002; **148**: 2831-2846.

80547. Johnson DR, Lee PKH, Holmes VF, Alvarez-Cohen L. An internal
806reference technique for accurately quantifying specific mRNAs by real-time
807PCR with application to the *tceA* reductive dehalogenase gene. Appl Environ
808Microbiol. 2005; **71**: 3866-3871.

80948. Yoon S, Cruz-Garcia C, Sanford R, Ritalahti KM, Loffler FE.
810Denitrification versus respiratory ammonification: environmental controls of
811two competing dissimilatory $NO3^{-}$/$NO2^{-}$ reduction pathways in Shewanella
812loihica strain PV-4. ISME J. 2015; **9**: 1093-1104.
81349. Sheu SY, Cho NT, Arun AB, Chen WM. Proposal of *Solimonas aquatica*
814sp. nov., reclassification of *Sinobacter flavus* Zhou *et al*. 2008 as *Solimonas*
815*flava* comb. nov. and *Singularimonas variicoloris* Friedrich and Lipski 2008 as
816*Solimonas variicoloris* comb. nov. and emended descriptions of the genus
817*Solimonas* and its type species *Solimonas soli*. Int J Syst Evol Microbiol. 2011;
818**61**: 2284-2291.
81950. Dubbels BL, Sayavedra-Soto LA, Bottomley PJ, Arp DJ. *Thauera*
820*butanivorans* sp. nov., a C2-C9 alkane-oxidizing bacterium previously
821referred to as '*Pseudomonas butanovora*'. Int J Syst Evol Microbiol. 2009; **59**:
8221576-1578.
82351. Yagi JM, Sims D, Brettin T, Bruce D, Madsen EL. The genome of
824*Polaromonas naphthalenivorans* strain CJ2, isolated from coal tar-
825contaminated sediment, reveals physiological and metabolic versatility and
826evolution through extensive horizontal gene transfer. Environ Microbiol.
8272009; **11**: 2253-2270.
82852. Finneran KT, Johnsen CV, Lovley DR. *Rhodoferax ferrireducens* sp. nov.,
829a psychrotolerant, facultatively anaerobic bacterium that oxidizes acetate
830with the reduction of Fe(III). Int J Syst Evol Microbiol. 2003; **53**: 669-673.

83153. Münch R, Hiller K, Grote A, Scheer M, Klein J, Schobert M, et al. Virtual
832Footprint and PRODORIC: an integrative framework for regulon prediction in
833prokaryotes. Bioinformatics. 2005; **15**: 4187-4189.
83454. Schmid MC, Hooper AB, Klotz MG, Woebken D, Lam P, Kuypers MM, et
835al. Environmental detection of octahaem cytochrome c
836hydroxylamine/hydrazine oxidoreductase genes of aerobic and anaerobic
837ammonium-oxidizing bacteria. Environ Microbiol. 2008; **10**: 3140-3149.
83855. Arp DJ, Chain PS, Klotz MG. The impact of genome analyses on our
839understanding of ammonia-oxidizing bacteria. Ann Rev Microbiol. 2007; **61**:
840503-528.
84156. Lau E, Fisher MC, Steudler PA, Cavanaugh CM. The methanol
842dehydrogenase gene, *mxaF*, as a functional and phylogenetic marker for
843proteobacterial methanotrophs in natural environments. PloS one. 2013; **8**:
844e56993.
84557. Chu F, Lidstrom ME. XoxF acts as the predominant methanol
846dehydrogenase in the type I methanotroph *Methylomicrobium buryatense*. J
847Bacteriol. 2016; **198**: 1317-1325.
84858. Kalyuzhnaya MG, Hristova KR, Lidstrom ME, Chistoserdova L.
849Characterization of a novel methanol dehydrogenase in representatives of
850*Burkholderiales*: implications for environmental detection of methylotrophy
851and evidence for convergent evolution. J Bacteriol. 2008; **190**: 3817-3823.
85259. Cooley RB, Dubbels BL, Sayavedra-Soto LA, Bottomley PJ, Arp DJ.
853Kinetic characterization of the soluble butane monooxygenase from *Thauera*

butanivorans, formerly '*Pseudomonas butanovora*.'. Microbiol. 2009; **155**: 2086-2096.

60. Holmes AJ. The diversity of soluble di-iron monooxygenases with bioremediation applications. In: Singh A, Kuhad R, Ward O, editors. Advances in Applied Bioremediation Soil Biology. Berlin: Springer; 2009.

61. Kotani T, Kawashima Y, Yurimoto H, Kato N, Sakai Y. Gene structure and regulation of alkane monooxygenases in propane-utilizing *Mycobacterium* sp. TY-6 and *Pseudonocardia* sp. TY-7. Journal of bioscience and bioengineering. 2006; **102**: 184-192.

62. Ashraf W, Mihdhir A, Murrell JC. Bacterial oxidation of propane. FEMS Microbiol Lett. 1994; **122**: 1-6.

63. Textor S, Wendisch VF, De Graaf AA, Müller U, Linder MI, Linder D, et al. Propionate oxidation in *Escherichia coli*: evidence for operation of a methylcitrate cycle in bacteria. Arch Microbiol. 1997; **168**: 428-436.

64. Suvorova IA, Ravcheev DA, Gelfand MS. Regulation and evolution of malonate and propionate catabolism in proteobacteria. J Bacteriol. 2012; **194:** 3234-3240.

65. Hausinger RP. New insights into acetone metabolism. J Bacteriol. 2007;189(3):671-3.

66. Kotani T, Yurimoto H, Kato N, Sakai Y. Novel acetone metabolism in a propane-utilizing bacterium, *Gordonia* sp. strain TY-5. J Bacteriol. 2007; **189**: 886-893.

87667. Willems A. The family *Comamonadaceae*. In: Rosenberg E, editor. The

877Prokaryotes - *Alphaproteobacteria* and *Betaproteobacteria*. Berlin: Springer;

8782014. p. 777-851.

879
880

881**Figure Legends**

882**Figure 1**. Maximum-likelihood tree of concatenated derived amino acid
883sequences of CuMMO-encoding genes *xmoC*, *xmoA* and *xmoB*. The tree was
884constructed as described in Materials & Methods. Preferred substrate or
885enzyme function, if known, for the CuMMOs are indicated in brackets.
886Accession numbers for sequences are given in Table S2. For the *Rhodoferax*
887and *Polaromonas* SAGs, the specific genomes encoding the CuMMOs are
888found in Table S3. For genomes encoding multiple CuMMOs, numerical
889identifiers were assigned to unique sequences (e.g. *xmoCAB1* or *xmoCAB2*).
890The scale bar represents substitutions per site. Branch support values are
891shown at each node and were determined based on 100 bootstrap replicates.
892

893**Figure 2.** Abundance of WIPMG *xmoA1* gene copies (per ml of water) during
894enrichment of oilsands tailings pond water under methane, ethane, propane,
895ammonium chloride or no added substrate. Error bars indicate ±1 SEM of
896triplicates.
897

898**Figure 3**. Abundance of WIPMG *xmoA1* gene copies (per ml) in SIP
899enrichments. Samples were enriched using isotopically light (^{12}C) or heavy
900(^{13}C) methane, ethane and propane or were left unamended. The bar graph
901indicates the number of *xmoA1* gene copies per SIP fraction. Error bars
902indicate ±1 SEM of two separate SIP gradients. The line graph indicates the

903relative DNA concentration per fraction with the highest quantity detected in
904any fraction set to 1.

905

906**Figure 4**. Gene arrangements of *xmoCABD* in SAGs: a – 17101 bp section of
907*Rhodoferax* SAG JGI 00BML02F20; b – 8526 bp section of *Rhodoferax* SAG JGI
90800BML02C18; c – 9822 bp section of *Polaromonas* SAG JGI 00BML02G21 d
909-.9887 bp segment of *Polaromonas* SAG JGI 00BML02L09. Sections (a) and (c)
910show the gene clusters corresponding to the WIPMG *xmoCAB1* cluster in
911Figure 1. Genes: *A* – *xmoA*, *B* – *xmoB*, *C* – *xmoC*, *D* – *xmoD*, 1 – long-chain
912fatty acid transport protein , 2 – poly (3-hyrdoxybutarate) depolymerase, 3 –
913aldehyde dehydrogenase family protein, 4 – cyclohexanecarboxylate-CoA
914ligase, 5 – class I SAM-dependent methyltransferase, 6 – 4-hydroxybutyrate
915CoA-transferase, 7 – hypothetical protein, 8 – DUF4242 domain-containing
916protein, 9 – copper resistance protein CopC, 10 – c-type cytochrome, 11 –
917transcriptional regulator, 12 – fatty-acid-CoA ligase, 13 – PQQ-dependent
918dehydrogenase, methanol/ethanol, 14 – acrylyl-CoA reductase (NADPH), 15 -
919competence protein ComEC. A single sigma-70 promoter (indicated by an
920arrow) was predicted in front of each *xmoCABD* operon via Virtual Footprint.

921

922**Figure 5**. Possible pathways for terminal (A) or sub-terminal (B) oxidation of
923propane in the *Rhodoferax* and *Polaromonas* SAGs. Both the citramalate (i)
924and methylcitrate (ii) pathways are shown in panel A. The propane

925monooxygenase could potentially include several enzymes, one of which

926may be a CuMMO.

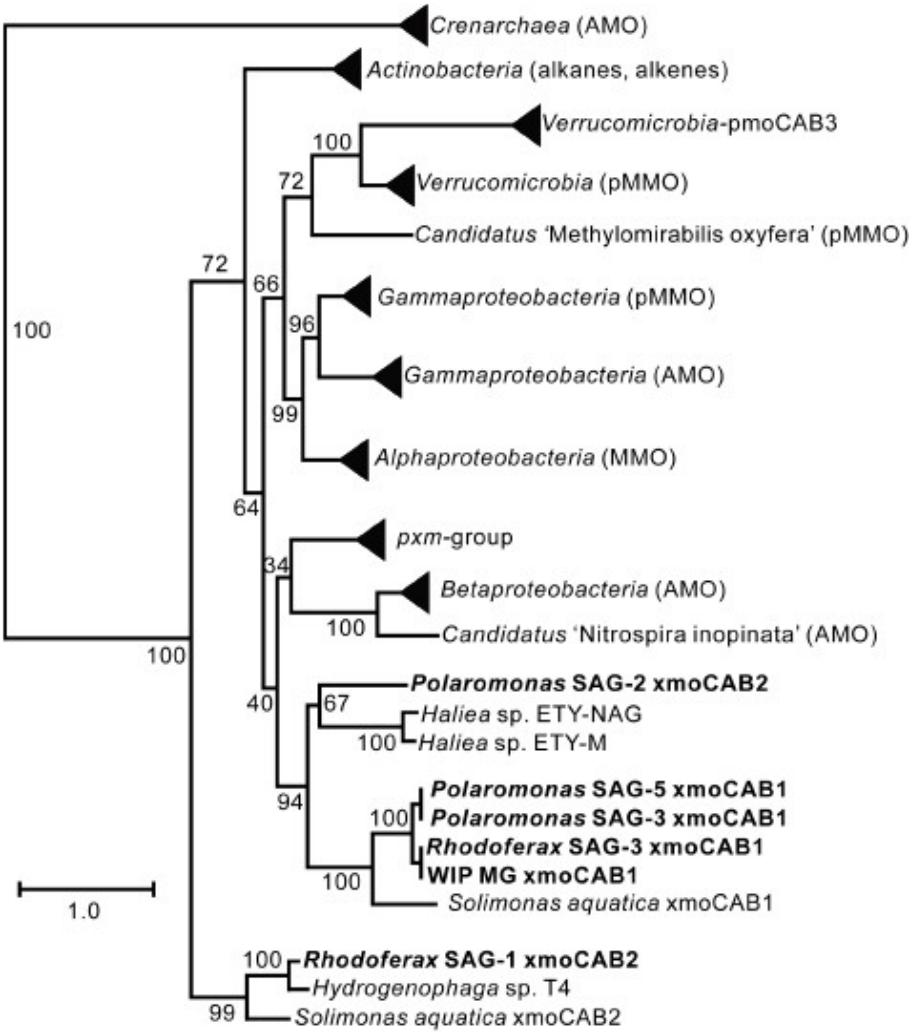
Table 1. *xmoA* gene and transcript copy numbers in the MLSB enrichment culture fed with a continuous supply of 8% v/v propane in air for 96 h and then air only afterwards. Specific assays for the WIPMG *xmoA1* gene cluster (targeting a group of related sequences from the metagenome and both SAG genera), as well as for the *xmoA2* gene of the *Rhodoferax* are shown. A primer set targeting the *xmoA2* cluster in the *Polaromonas* failed to amplify a product.

Time (h)	WIPMG *xmoA1* ($\times 10^4$ copies per ml)[a]			*Rhodoferax xmoA2* ($\times 10^4$ copies per ml)[a]		
	Gene	Transcript	Transcripts per gene [b]	Gene	Transcript	Transcripts per gene [b]
48 (+C_3H_8)	15 (1.2)	3.8 (0.84)	0.25 (0.06)	10 (8.4)	2.2 (1.1)	0.22 (0.21)
96 (+C_3H_8)	390 (6.0)	160 (7.9)	0.42 (0.07)	32 (3.9)	22 (3.7)	0.70 (0.15)
120 (-C_3H_8)	340 (28)	4.7 (1.0)	0.01 (0.00)	16 (5.4)	2.6 (1.4)	0.17 (0.10)

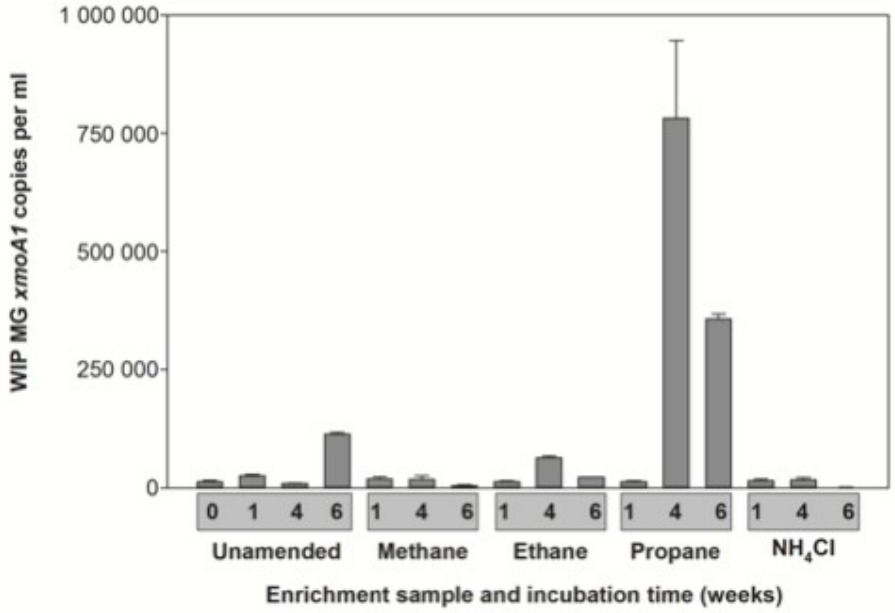
[a] ± 1 standard deviation of technical triplicate samples processed through extraction and purification procedures in parallel (in parentheses)

[b] The standard deviations were calculated using the propagation of error method

Figure 1. Maximum-likelihood tree of concatenated derived amino acid sequences of CuMMO-encoding genes *xmoC*, *xmoA* and *xmoB*. The tree was constructed as described in Materials & Methods. Preferred substrate or enzyme function, if known, for the CuMMOs are indicated in brackets. Accession numbers for sequences are given in Table S2. For the *Rhodoferax* and *Polaromonas* SAGs, the specific genomes encoding the CuMMOs are found in Tables S3. For genomes encoding multiple CuMMOs, numerical identifiers were assigned to unique sequences (e.g. *xmoCAB1* or *xmoCAB2*). The scale bar represents 1.0 substitution per site. Branch support values are shown at each node and were determined based on 100 bootstrap replicates.



962**Figure 2.** Abundance of WIPMG *xmoA1* gene copies (per ml of water) during
963enrichment of oilsands tailings pond water under methane, ethane, propane,
964ammonium chloride or no added substrate. Error bars indicate ±1 SEM of
965triplicates.

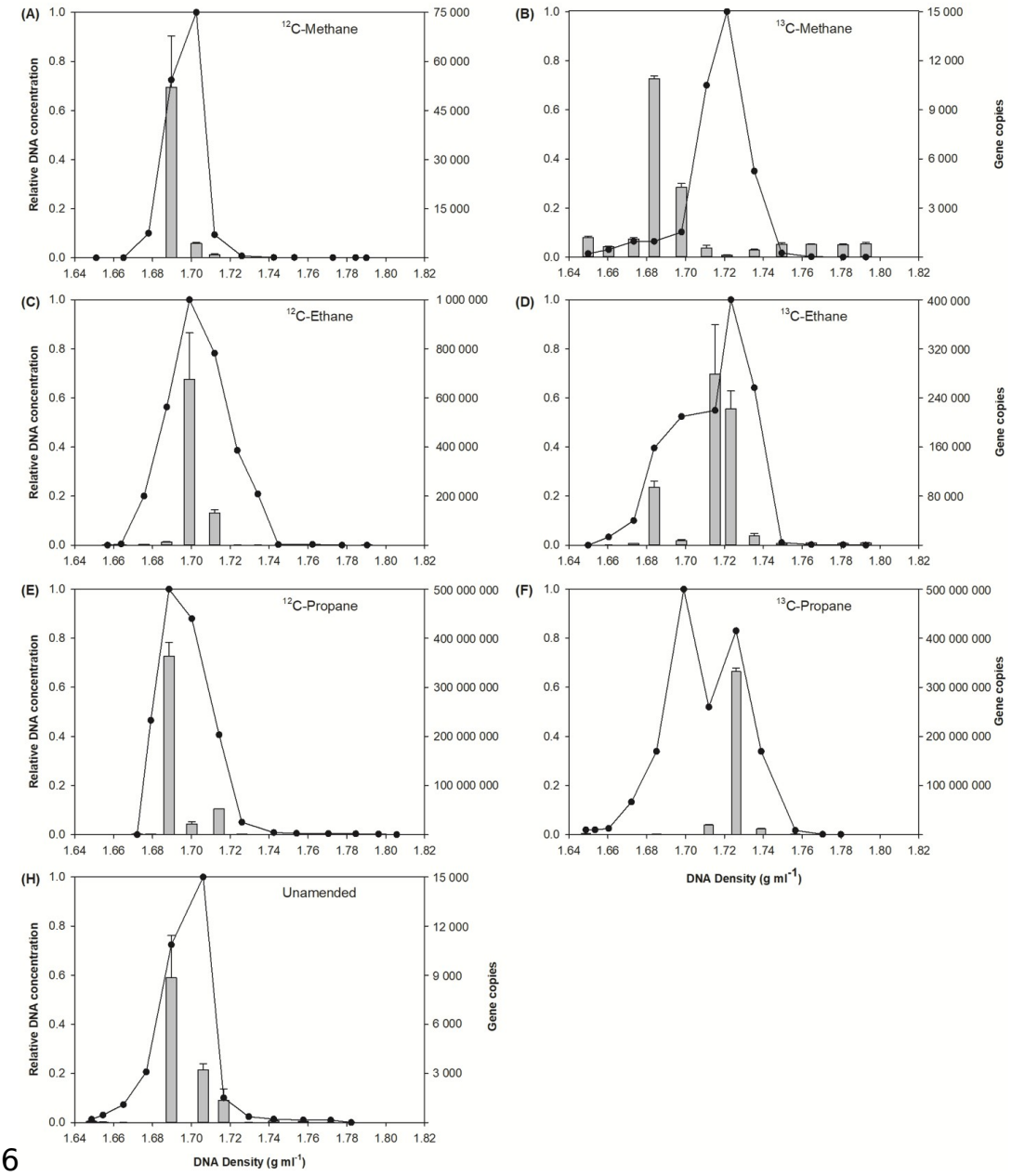


966

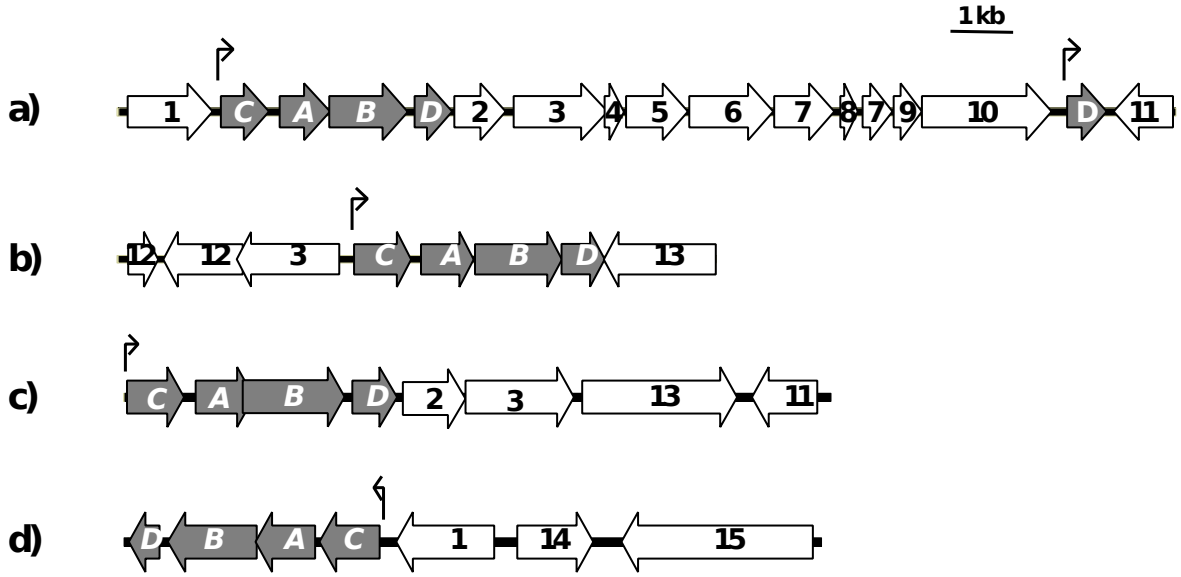
97 49
98

967
968

Figure 3. Abundance of WIPMG *xmoA1* gene copies (per ml) in SIP enrichments. Samples were enriched using isotopically light (^{12}C) or heavy (^{13}C) methane, ethane and propane or were left unamended. The bar graph indicates the number of *xmoA1* gene copies per SIP fraction. Error bars indicate ±1 SEM of two separate SIP gradients. The line graph indicates the relative DNA concentration per fraction with the highest quantity detected in any fraction set to 1.


(A)
^{12}C-Methane
(B)
^{13}C-Methane
(C)
^{12}C-Ethane
(D)
^{13}C-Ethane
(E)
^{12}C-Propane
(F)
^{13}C-Propane
(H)
Unamended
Relative DNA concentration
Gene copies
DNA Density (g ml^{-1})

978**Figure 4**. Gene arrangements of *xmoCABD* in SAGs: a – 17101 bp section of
979*Rhodoferax* SAG JGI 00BML02F20; b – 8526 bp section of *Rhodoferax* SAG JGI
98000BML02C18; c – 9822 bp section of *Polaromonas* SAG JGI 00BML02G21 d
981-.9887 bp segment of *Polaromonas* SAG JGI 00BML02L09. Sections a and d
982show the gene clusters corresponding to the WIPMG *xmoCAB1* cluster in
983Figure 1. Genes: *A* – *xmoA*, *B* – *xmoB*, *C* – *xmoC*, *D* – *xmoD*, 1 – long-chain
984fatty acid transport protein , 2 – poly (3-hyrdoxybutarate) depolymerase, 3 –
985aldehyde dehydrogenase family protein, 4 – cyclohexanecarboxylate-CoA
986ligase, 5 – class I SAM-dependent methyltransferase, 6 – 4-hydroxybutyrate
987CoA-transferase, 7 – hypothetical protein, 8 – DUF4242 domain-containing
988protein, 9 – copper resistance protein CopC, 10 – c-type cytochrome, 11 –
989transcriptional regulator, 12 – fatty-acid-CoA ligase, 13 – PQQ-dependent
990dehydrogenase, methanol/ethanol, 14 – acrylyl-CoA reductase (NADPH), 15 -
991competence protein ComEC. A single promoter (indicated by an arrow) was
992predicted in front of each *xmoCABD* operon via Virtual Footprint.


1kb
a)
1
C
A
B
D
2
3
4
5
6
7
8
7
9
10
D
11
b)
12
12
3
C
A
B
D
13
c)
C
A
B
D
2
3
13
11
d)
D
B
A
C
1
14
15

993

994

Figure 5. Possible pathways for terminal (A) or sub-terminal (B) oxidation of propane in the *Rhodoferax* and *Polaromonas* SAGs. Both the citramalate (i) and methylcitrate (ii) pathways are shown in panel A. The propane monooxygenase could potentially include several enzymes, one of which may be a CuMMO.